

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-219830

(43) 公開日 平成7年(1995)8月18日

(51) Int.Cl.⁴

G 0 6 F 12/00

識別記号

5 4 5 A 8944-5B

庁内整理番号

F I

技術表示箇所

審査請求 未請求 請求項の数29 O L (全 16 頁)

(21) 出願番号 特願平7-3453

(22) 出願日 平成7年(1995)1月12日

(31) 優先権主張番号 08/181704

(32) 優先日 1994年1月14日

(33) 優先権主張国 米国 (US)

(71) 出願人 391055933

マイクロソフト コーポレイション

MICROSOFT CORPORATI
ON

アメリカ合衆国 ワシントン州 98052-

6399 レッドモンド ワン マイクロソフ
ト ウェイ (番地なし)

(72) 発明者 ユーヴァル ニーマン

アメリカ合衆国 ワシントン州 98004

ベルヴィュー ナインティファースト プ
レイス ノースイースト 2765

(74) 代理人 弁理士 中村 稔 (外6名)

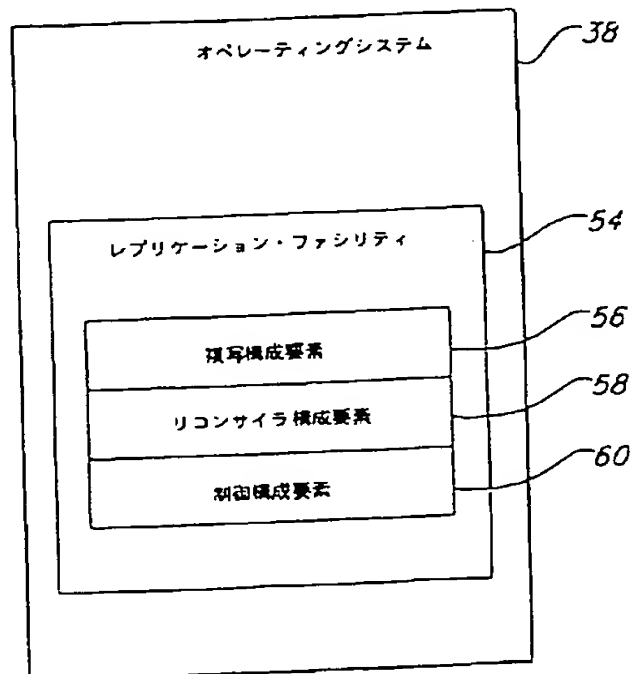
最終頁に続く

(54) 【発明の名称】 レプリケーション・ファシリティ

(57) 【要約】

【目的】 分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給するレプリケーション・ファシリティを提供する。

【構成】 本発明の方法は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含む多数のコンピュータ・システムを有している分散型システムにおいて、ファイルのツリーに編成された複数のファイルを供給し、ファイルのコピーが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて一つのコンピュータ・システムの記憶装置に記憶される単一のファイルをレプリケートし、ファイルのサブツリーのコピーが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて一つのコンピュータ・システムの記憶装置に記憶される、ファイルのツリーから、多重段階のファイルのサブツリーをレプリケートする段階を具備する。



【特許請求の範囲】

【請求項1】 レプリケーション・ファシリティ及びそれぞれが記憶装置を含む多数のコンピュータ・システムを有している分散型システムにおいて、ファイルのツリーに編成された複数のファイルを供給し、前記ファイルのコピーが別の前記コンピュータ・システムの記憶装置に記憶されるように前記レプリケーション・ファシリティを用いて一つの前記コンピュータ・システムの記憶装置に記憶される単一の前記ファイルをレプリケートし、ファイルのサブツリーのコピーが別の前記コンピュータ・システムの前記記憶装置に記憶されるように前記レプリケーション・ファシリティを用いて一つの前記コンピュータ・システムの前記記憶装置に記憶される、前記ファイルのツリーから、多重段階のファイルの前記サブツリーをレプリケートする段階を具備することを特徴とする方法。

【請求項2】 前記ファイルのコピーが追加の前記コンピュータ・システムの記憶装置に記憶されるように前記レプリケーション・ファシリティを用いて前記単一ファイルをレプリケートする段階を更に具備することを特徴とする請求項1に記載の方法。

【請求項3】 前記サブツリーのコピーが前記追加のコンピュータ・システムの前記記憶装置に記憶されるように前記レプリケーション・ファシリティを用いて前記サブツリーをレプリケートする段階を更に具備することを特徴とする請求項1に記載の方法。

【請求項4】 前記レプリケートされているサブツリーは、少なくとも三段階のファイルを含むことを特徴とする請求項3に記載の方法。

【請求項5】 それぞれがファイルを記憶する記憶装置を含んでいる複数のコンピュータ・システムと、前記ファイルのネームのツリー構造として前記システムのネーム空間を管理するネーム空間管理プログラムと、前記ネーム空間のサブツリーを含んでいるあらゆるファイルをレプリケートするレプリケーション・ファシリティとを備えていることを特徴とする分散型システム。

【請求項6】 リコンサイラ・ファシリティ及び多数のコンピュータ・システムを有している分散型システムにおいて、一つの前記コンピュータ・システムのファイルの第1のコピー及び別の該コンピュータ・システムの前記ファイルの第2のコピーを供給し、最後に一致されて以来前記ファイルの前記第1のコピーに対してなされたあらゆる変更を前記ファイルの前記第2のコピーが組み込むように前記リコンサイラ・ファシリティを用いて該ファイルの該第1のコピーを該ファイルの該第2のコピーと一致させ、一つの前記コンピュータ・システムの一

群のファイルの第1のコピー及び別の該コンピュータ・システムの一群のファイルの第2のコピーを供給し、最後に一致されて以来前記一群のファイルの前記第1のコピーに対してなされたあらゆる変更を前記一群のファイルの前記第2のコピーが組み込むように前記リコンサイラ・ファシリティを用いて該一群のファイルの該第1のコピーを該一群のファイルの該第2のコピーと一致させる段階を具備することを特徴とする方法。

【請求項7】 前記一群のファイルの前記第1のコピーを前記一群のファイルの前記第2のコピーと一致させる段階は、該一群のファイルの該第1のコピーの各ファイルを該一群のファイルの該第2のコピーの対応しているファイルと対基準により一対で一致させる段階を更に含んでいることを特徴とする請求項6に記載の方法。

【請求項8】 レプリケーション・ファシリティ及びそれぞれが記憶装置を含んでいる多数のコンピュータ・システムを有している分散型システムにおいて、第1の前記コンピュータ・システムの前記記憶装置に記憶される一群のファイルの第1のコピーを供給し、第2の前記コンピュータ・システムの前記記憶装置に記憶される前記一群のファイルの第2のコピーを供給し、前記一群のファイルの前記第1のコピーの前記ファイルの少なくとも一つに対して変更を行い、事象の発生により前記一群のファイルの前記第2のコピーへ前記変更を伝搬し、前記一群のファイルの前記第1のコピーの前記ファイルの少なくとも一つに対して追加の変更を行い、別の事象の発生により前記一群のファイルの前記第2のコピーへ前記追加の変更を伝搬する段階を具備することを特徴とする方法。

【請求項9】 前記事象は、所定の時限の経過であることを特徴とする請求項8に記載の方法。

【請求項10】 前記別の事象も、所定の時限の経過であることを特徴とする請求項8に記載の方法。

【請求項11】 前記事象は、前記変更を受け取るための前記第2のコンピュータ・システムによる要求であることを特徴とする請求項8に記載の方法。

【請求項12】 前記別の事象は、前記追加の変更を受け取るための前記第2のコンピュータ・システムによる要求であることを特徴とする請求項8に記載の方法。

【請求項13】 前記一群のファイルの前記第2のコピーが前記一群のファイルの前記第1のコピーに対してなされた前記変更を組み込むように該一群のファイルの該第2のコピーを該一群のファイルの該第1のコピーと一致させる段階を更に含むことを特徴とする請求項8に記載の方法。

【請求項14】 前記一群のファイルの前記第2のコピーが前記一群のファイルの前記第1のコピーに対してなされた前記追加の変更を組み込むように該一群のファイルの該第2のコピーを該一群のファイルの該第1のコピーと一致させる段階を更に含むことを特徴とする請求項13に記載の方法。

【請求項15】 レプリケーション・ファシリティ及びそれぞれが記憶装置を含むコンピュータ・システムを有している分散型システムにおいて、前記コンピュータ・

システムの前記記憶装置に、ネームを有している、ファイルを記憶し、前記記憶されたファイルの前記ネームの論理的編成からなる分散型ネーム空間を供給し、前記ファイルの選択された部分を保持している新しいファイルを生成するために前記レプリケーション・ファシリティを用いてそのネームが前記分散型ネーム空間の一部を形成しかつ一つの前記コンピュータ・システムの前記記憶装置に記憶された一群のファイルの選択された部分をレプリケートする段階を具備することを特徴とする方法。

【請求項 16】 前記分散型システムの前記コンピュータ・システムの少なくとも部分にわたり前記新しいファイルを分散すべく前記新しいファイルをレプリケートする段階を更に具備することを特徴とする請求項 15 に記載の方法。

【請求項 17】 第 1 のコンピュータ・システム及び第 2 のコンピュータ・システムを有している分散型システムにおいて、前記第 1 のコンピュータ・システムに記憶される所与のクラスの一組のファイルの第 1 のコピーを供給し、前記第 2 のコンピュータ・システムに記憶される前記所与のクラスの前記一組のファイルの第 2 のコピーを供給し、前記所与のクラスのファイルだけを一致させるクラス特定リコンサイラを用いて前記一組のファイルの前記第 1 のコピーを前記一組のファイルの前記第 2 のコピーと一致させる段階を具備することを特徴とする方法。

【請求項 18】 前記一組のファイルの前記第 1 のコピーに対して変更を行い、クラスに関係なくファイルを一致させるクラス独立リコンサイラを用いて前記一組のファイルの前記第 1 のコピーを前記一組のファイルの前記第 2 のコピーと一致させる段階を更に含むことを特徴とする請求項 17 に記載の方法。

【請求項 19】 専用レプリケーション機構及びそれぞれが記憶装置を含んでおり処理を実行するためのコンピュータ・システムを有している分散型システムにおいて、一つの前記コンピュータ・システム上でアプリケーション・プログラムを実行し、前記アプリケーション内にてそれぞれがファイルをアクセスすることを許容される処理のリストを維持している一組のファイルをリプリケートすべく前記専用レプリケーション機構に対して要求を行い、前記ファイルをアクセスすることを許容される前記処理のリストを、各ファイルに対して、レプリケートすることなく新しい組のファイルを生成すべく前記専用レプリケーション機構を用いて前記一組のファイルをレプリケートする段階を具備することを特徴とする方法。

【請求項 20】 第 1 のコンピュータ・システム及び第 2 のコンピュータ・システムを有している分散型システムにおいて、前記第 1 のコンピュータ・システムでファイルのコレクションを供給し、前記ファイルのコレクションを前記第 2 のコンピュータ・システムへレプリケー

トするための要求に応じて、前記コレクションの全ての前記ファイルがレプリケートされるべきかまたはされるべきでないかを決定し、前記コレクションの全ての前記ファイルがレプリケートされるべきであると決定されるならば、前記コレクションのレプリカが前記第 2 のコンピュータ・システムで供給されるように該コレクションの全ての該ファイルをレプリケートし、前記コレクションの全ての前記ファイルがレプリケートされるべきでないとして決定されるならば、前記コレクションの前記ファイルの全てをレプリケートしない段階を具備することを特徴とする方法。

【請求項 21】 第 1 のコンピュータ・システム及び第 2 のコンピュータ・システムを有している分散型システムにおいて、前記第 1 のコンピュータ・システムの一組のファイルの第 1 のコピーを供給し、前記第 2 のコンピュータ・システムの前記一群のファイルの第 2 のコピーを供給し、前記一群のファイルの前記第 1 のコピーに対して変更を行い、それぞれが前記一群のファイルの前記第 1 のコピーの前記ファイルをアクセスしかつ読取るためのアクセス権を有するエージェントを該一群のファイルの該第 1 のコピーに対して供給し、前記一群のファイルの前記第 2 のコピーを前記一群のファイルの前記第 1 のコピーと一致させるために前記第 2 のコンピュータ・システムでリコンサイラを供給し、前記一群のファイルの前記第 1 のコピーの前記ファイルをアクセスしかつ読取るべく限定された権限を前記リコンサイラに賦与するプロキシを、該一群のファイルの該第 1 のコピーの前記エージェントから該リコンサイラへ賦与し、前記一群のファイルの前記第 1 のコピーに対してなされた前記変更が前記一群のファイルの前記第 2 のコピーに対してなされるように前記リコンサイラを用いて該一群のファイルの該第 2 のコピーを該一群のファイルの該第 1 のコピーと一致させる段階を具備することを特徴とする方法。

【請求項 22】 分散型システムにおいて、前記分散型システムの非均質ファイル・システムを供給し、前記ファイル・システムのファイルへのアクセスを管理すべく各ファイル・システムに対して記憶管理プログラムを供給し、第 1 の組のファイルを第 2 の組のファイルと一致させるための要求に応じて、該第 1 の組のファイルを保持する前記ファイル・システムに対する前記記憶管理プログラムにより該第 1 の組のファイルへのアクセスを賦与し、かつ該第 2 の組のファイルを保持する前記ファイル・システムに対する前記記憶管理プログラムにより該第 2 の組のファイルへのアクセスを賦与し、前記第 1 の組のファイル及び前記第 2 の組のファイルを保持しているそれぞれのファイル・システムの前記記憶管理プログラムの制御の下で第 1 のオブジェクト・セットを第 2 の組のオブジェクトと一致させる段階を具備することを特徴とする方法。

【請求項 23】 前記ファイル・システムに記憶される

ファイルの各コピーは、前記記憶管理プログラムによって記憶特定識別子が供給されることを特徴とする請求項22に記載の方法。

【請求項24】 各記憶管理プログラムは、そのファイル・システムの前記ファイルへ変更を報告することを特徴とする請求項22に記載の方法。

【請求項25】 前記変更は、ファイルの消去を含むことを特徴とする請求項24に記載の方法。

【請求項26】 前記変更は、ファイルの改名を含むことを特徴とする請求項24に記載の方法。

【請求項27】 前記変更は、前記分散型システムのファイルの移動を含むことを特徴とする請求項24に記載の方法。

【請求項28】 前記変更は、変更記録簿へ報告されかつ前記一致させる段階は、該変更記録簿を用いて実行されることを特徴とする請求項24に記載の方法。

【請求項29】 ファイルの各コピーは、固有の識別子に割り当てられかつ前記一致させる段階は、どのファイルが一致させられるのかを決定するために識別子と比較することを含むことを特徴とする請求項22に記載の方法。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、データ処理システムに関し、より特定のには、分散型システム内で用いられるレプリケーション・ファシリティ (replication facilities) に関する。

【0002】

【従来の技術】 レプリケーション・ファシリティは、多数の異なる形式のソフトウェア・プロダクトに供給されている。例えば、レプリケーション・ファシリティは、データベース・プロダクト (database products)、ネットワーク・ディレクトリー・サービス・プロダクト (network directory service products)、及びグループウェア・プロダクト (groupware products) に組み込まれている。

【0003】

【発明が解決しようとする課題】 しかしながら、通常のレプリケーション・ファシリティの多くは、何をそれらがレプリケート (replicate) できるかについて制限される。例えば、多くの通常のレプリケートは、一形式の論理構造 (例えば、ファイル) をレプリケートできるだけである。更に、通常のレプリケートは、一度にレプリケートされる論理的構造の量について制限される。特に、多くの通常のレプリケートは、一度に一つのファイルだけをレプリケートできる。本発明の目的は、上記従来のレプリケーション・ファシリティにおける問題点に鑑み、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給するレプリケーション・ファシリティを提供することである。本発明の他の目的

は、分散型環境の分散型ネーム空間内であらゆるサブツリーをレプリケートすることができるレプリケーション・ファシリティを提供することである。本発明の別の目的は、多重マスターの、弱く一貫するレプリケーションを供給するレプリケーション・ファシリティを提供することである。

【0004】 本発明の更なる目的は、公開レプリケーションと専用レプリケーションの両方を支援するレプリケーション・ファシリティを提供することである。

10 【0005】

【課題を解決するための手段】 本発明の上記目的は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含む多数のコンピュータ・システムを有している分散型システムにおいて、ファイルのツリーに編成された複数のファイルを供給し、ファイルのコピーが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて一つのコンピュータ・システムの記憶装置に記憶される単一のファイルをレプリケートし、ファイルのサブツリーのコピーが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて一つのコンピュータ・システムの記憶装置に記憶される、ファイルのツリーから、多重段階のファイルのサブツリーをレプリケートする段階を具備する方法によって達成される。本発明の方法において、ファイルのコピーが追加のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて単一ファイルをレプリケートする段階を更に具備するように構成してもよい。本発明の方法において、サブツリーのコピーが追加のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いてサブツリーをレプリケートする段階を更に具備するように構成してもよい。

【0006】 本発明の方法において、レプリケートされているサブツリーは、少なくとも三段階のファイルを含むように構成してもよい。また、上述した本発明の目的は、それぞれがファイルを記憶する記憶装置を含んでいる複数のコンピュータ・システムと、ファイルのネームのツリー構造としてシステムのネーム空間を管理するネーム空間管理プログラムと、ネーム空間のサブツリーを含んでいるあらゆるファイルをレプリケートするレプリケーション・ファシリティとを備えている分散型システムによっても達成される。更に、上述した本発明の目的は、リコンティラ・ファシリティ及び多数のコンピュータ・システムを有している分散型システムにおいて、一つのコンピュータ・システムのファイルの第1のコピー及び別のコンピュータ・システムのファイルの第2のコピーを供給し、最後に一致されて以来ファイルの第1のコピーに対してなされたあらゆる変更をファイルの第2のコピーが組み込むようにリコンサイラ・ファシリティ

を用いてファイルの第1のコピーをファイルの第2のコピーと一致させ、一つのコンピュータ・システムの一部のファイルの第1のコピー及び別のコンピュータ・システムの一部のファイルの第2のコピーを供給し、最後に一致されて以来一群のファイルの第1のコピーに対してなされたあらゆる変更を一群のファイルの第2のコピーが組み込むようにリコンサイラ・ファシリティを用いて一群のファイルの第1のコピーを一群のファイルの第2のコピーと一致させる段階を具備する方法によっても達成される。

【0007】本発明において、一群のファイルの第1のコピーを一群のファイルの第2のコピーと一致させる段階は、一群のファイルの第1のコピーの各ファイルを一群のファイルの第2のコピーの対応しているファイルと対基準により一対で一致させる段階を更に含むように構成してもよい。本発明の上述した目的は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含んでいる多数のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムの記憶装置に記憶される一群のファイルの第1のコピーを供給し、第2のコンピュータ・システムの記憶装置に記憶される一群のファイルの第2のコピーを供給し、一群のファイルの第1のコピーのファイルの少なくとも一つに対して変更を行い、事象の発生により一群のファイルの第2のコピーへ変更を伝搬し、一群のファイルの第1のコピーのファイルの少なくとも一つに対して追加の変更を行い、別の事象の発生により一群のファイルの第2のコピーへ追加の変更を伝搬する段階を具備する方法によっても達成される。本発明において、事象は、所定の時限の経過であるように構成してもよい。

【0008】本発明において、別の事象も、所定の時限の経過であるように構成してもよい。本発明において、事象は、変更を受け取るための第2のコンピュータ・システムによる要求であるように構成してもよい。本発明において、別の事象は、追加の変更を受け取るための第2のコンピュータ・システムによる要求であるように構成してもよい。本発明において、一群のファイルの第2のコピーが一群のファイルの第1のコピーに対してなされた変更を組み込むように一群のファイルの第2のコピーを一群のファイルの第1のコピーと一致させる段階を更に含むように構成してもよい。本発明において、一群のファイルの第2のコピーが一群のファイルの第1のコピーに対してなされた追加の変更を組み込むように一群のファイルの第2のコピーを一群のファイルの第1のコピーと一致させる段階を更に含むように構成してもよい。上述した本発明の目的は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含むコンピュータ・システムを有している分散型システムにおいて、コンピュータ・システムの記憶装置に、ネームを有している、ファイルを記憶し、記憶されたファイルのネームの論理

的編成からなる分散型ネーム空間を供給し、ファイルの選択された部分を保持している新しいファイルを生成するためにレプリケーション・ファシリティを用いてそのネームが分散型ネーム空間の一部を形成しかつ一つのコンピュータ・システムの記憶装置に記憶された一群のファイルの選択された部分をレプリケートする段階を具備する方法によっても達成される。

【0009】本発明において、分散型システムのコンピュータ・システムの少なくとも部分にわたり新しいファイルを分散すべく新しいファイルをレプリケートする段階を更に具備するように構成してもよい。また、上記本発明の目的は、第1のコンピュータ・システム及び第2のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムに記憶される所与のクラスの一組のファイルの第1のコピーを供給し、第2のコンピュータ・システムに記憶される所与のクラスの一組のファイルの第2のコピーを供給し、所与のクラスのファイルだけを一致させるクラス特定リコンサイラを用いて一組のファイルの第1のコピーを一組のファイルの第2のコピーと一致させる段階を具備する方法によっても達成される。本発明において、一組のファイルの第1のコピーに対して変更を行い、クラスに関係なくファイルを一致させるクラス独立リコンサイラを用いて一組のファイルの第1のコピーを一組のファイルの第2のコピーと一致させる段階を更に含むように構成してもよい。更に、本発明の上記目的は、専用レプリケーション機構及びそれぞれが記憶装置を含んでおり処理を実行するためのコンピュータ・システムを有している分散型システムにおいて、一つのコンピュータ・システム上でアプリケーション・プログラムを実行し、アプリケーション内にてそれぞれがファイルをアクセスすることを許容される処理のリストを維持している一組のファイルをレプリケートすべく専用レプリケーション機構に対して要求を行い、ファイルをアクセスすることを許容される処理のリストを、各ファイルに対して、レプリケートすることなく新しい組のファイルを生じすべく専用レプリケーション機構を用いて一組のファイルをレプリケートする段階を具備する方法によっても達成される。

【0010】本発明の上記目的は、第1のコンピュータ・システム及び第2のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムでファイルのコレクションを供給し、ファイルのコレクションを第2のコンピュータ・システムへレプリケートするための要求に応じて、コレクションの全てのファイルがレプリケートされるべきかまたははされるべきでないかを決定し、コレクションの全てのファイルがレプリケートされるべきであると決定されるならば、コレクションのレプリカが第2のコンピュータ・システムで供給されるようにコレクションの全てのファイルをレプリケートし、コレクションの全てのファイルがレプリケ

ートされるべきでないとして決定されるならば、コレクションのファイルの全てをレプリケートしない段階を具備する方法によっても達成される。また、本発明の上述した目的は、第1のコンピュータ・システム及び第2のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムの一組のファイルの第1のコピーを供給し、第2のコンピュータ・システムの一組のファイルの第2のコピーを供給し、一組のファイルの第1のコピーに対して変更を行い、それぞれが一組のファイルの第1のコピーのファイルにアクセスしかつ読取るためのアクセス権を有するエージェントを一組のファイルの第1のコピーに対して供給し、一組のファイルの第2のコピーを一組のファイルの第1のコピーと一致させるために第2のコンピュータ・システムでリコンサイラを供給し、一組のファイルの第1のコピーのファイルにアクセスしかつ読取るべく限定された権限をリコンサイラに賦与するプロセスを、一組のファイルの第1のコピーのエージェントからリコンサイラへ賦与し、一組のファイルの第1のコピーに対してなされた変更が一組のファイルの第2のコピーに対してなされるようにリコンサイラを用いて一組のファイルの第2のコピーを一組のファイルの第1のコピーと一致させる段階を具備する方法によっても達成される。

【0011】更に、本発明の目的は、分散型システムにおいて、分散型システムの非均質ファイル・システムを供給し、ファイル・システムのファイルへのアクセスを管理すべく各ファイル・システムに対して記憶管理プログラムを供給し、第1の組のファイルを第2の組のファイルと一致させるための要求に応じて、第1の組のファイルを保持するファイル・システムに対する記憶管理プログラムにより第1の組のファイルへのアクセスを賦与し、かつ第2の組のファイルを保持するファイル・システムに対する記憶管理プログラムにより第2の組のファイルへのアクセスを賦与し、第1の組のファイル及び第2の組のファイルを保持しているそれぞれのファイル・システムの記憶管理プログラムの制御の下で第1のオブジェクト・セットを第2の組のオブジェクトと一致させる段階を具備する方法によっても達成される。本発明において、ファイル・システムに記憶されるファイルの各コピーは、記憶管理プログラムによって記憶特定識別子が供給されるように構成してもよい。本発明において、各記憶管理プログラムは、そのファイル・システムのファイルへ変更を報告するように構成してもよい。

【0012】本発明において、変更は、ファイルの消去を含むように構成してもよい。本発明において、変更は、ファイルの改名を含むように構成してもよい。本発明において、変更は、分散型システムのファイルの移動を含むように構成してもよい。本発明において、変更は、変更記録簿へ報告されかつ一致させる段階は、変更記録簿を用いて実行されるように構成してもよい。本発

明において、ファイルの各コピーは、固有の識別子に割り当てられかつ一致させる段階は、どのファイルが一致させられるのかを決定するために識別子と比較することを含むように構成してもよい。

【0013】

【作用】本発明の好ましい実施例の第1の観点によれば、レプリケーション・ファシリティ及びそれぞれが記憶装置を含む多数のコンピュータ・システムを有している分散型システムにおいて方法が実行される。この方法では、複数のファイルが供給されかつツリー（木）に編成される。ファイルの単一のものは、ファイルのオリジナル・コピーよりも、ファイルのコピーが異なるコンピュータシステムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いてレプリケートされる。多重レベルのファイルのサブツリーもレプリケートされる。サブツリーは、一つのコンピュータ・システムの記憶装置に初め記憶されている。レプリケーションは、サブツリーのコピー及びそのファイルが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて実行される。本発明の別の観点によれば、ファイルの第1のコピーは、一つのコンピュータ・システムに供給される。ファイルの第2のコピーは、別のコンピュータ・システムに供給される。ファイルの第1のコピーは、リコンサイラ・ファシリティ（reconciler facility）を用いてファイルの第2のコピーと一致される。リコンシリエーション（reconciliation）は、ファイルの第2のコピーがファイルの第1のコピーになされたあらゆる変更（changes）を含むことを確実にする。一組のファイルの第1のコピーは、一つのコンピュータ・システムに供給され、そして一組のファイルの第2のコピーは、別のコンピュータ・システムに供給される。リコンサイラ・ファシリティは、最後に一致されていらい一組のファイルの第2のコピーが一組のファイルの第1のコピーになされたあらゆる変更を含むように一組のファイルの第1のコピーを一組のファイルの第2のコピーと一致させるべく用いられる。

【0014】本発明の更なる観点によれば、一組のファイルの第1のコピーは、第1のコンピュータ・システムの記憶装置に記憶される。一組のファイルの第2のコピーは、第2のコンピュータ・システムの記憶装置に記憶される。変更は、一組のファイルの第1のコピーのファイルの少なくとも一つに対してなされる。変更は、事象の発生によりファイルの第2の群へ伝達される。追加の変更は、一組のファイルの第1のコピーのファイルの少なくとも一つに対してなされ、そしてこれらの変更は、別の事象の発生により一組のファイルの第2のコピーへも伝達される。本発明の別の観点によれば、一組のファイルの第1のコピーは、第1のコンピュータ・システムの記憶装置に記憶される。一組のファイルの第2のコピーは、第2のコンピュータ・システムの記憶装置に記憶

される。一群のファイルの第1のコピーに対してなされたあらゆる変更は、変更が一群のファイルの第2のコピーに対してなされうように第2のコンピュータ・システムへ増加的に送られる。本発明の追加の観点によれば、記憶装置の一つに記憶されるファイルの第1のセットは、レプリケートされるべく特定される。フィルタは、ファイルの第1のセットのどのファイルがレプリケートされるべきかを決定するために特定される。フィルタによって特定されたファイルは、ファイルの第2のセットを生成すべくレプリケーション・ファシリティを用いてレプリケートされる。

【0015】本発明の更なる観点によれば、ネームを有しているファイルは、分散型システムのコンピュータ・システムの記憶装置に記憶される。分散型ネーム空間は、供給される。分散型ネーム空間は、記憶されたファイルのネームの論理的編成 (logical organization) を含む。ネーム空間の一群のファイル選択された部分は、ファイルの選択された部分を保持している新しいファイルを生成すべくレプリケートされる。本発明の更なる観点によれば、所与クラスの一セットのファイルの第1のコピーは、第1のコンピュータ・システムに記憶される。一セットのファイルの第2のコピーは、第2のコンピュータ・システムに記憶される。一セットのファイルの第1のコピーは、所与クラスのファイルだけを一致するクラス特定リコンサイラ (class-specific reconciler) を用いて一セットの第2のコピーと一致される。ファイルは、クラスに編成される、持続オブジェクト (persistent objects) として記憶されう。オブジェクト及びクラスは、後述する。本発明の別の観点によれば、アプリケーション・プログラムは、分散型システムのコンピュータ・システムの一つで実行される。一セットのファイルをレプリケートすべく、要求がアプリケーション・プログラム内で専用レプリケーション機構に対してなされる。ファイルのそれぞれは、ファイルにアクセスすることが許される処理の一覧表を維持する。一セットのファイルは、ファイルにアクセスすることが許される処理の一覧表をレプリケートしないでファイルの新しいセットを生成すべく専用レプリケーション機構 (private replication mechanism) を用いてレプリケートされる。

【0016】本発明の更なる観点によれば、一群のファイルの第1のコピーは、第1のコンピュータ・システムに供給され、そして一群のファイルの第2のコピーは、第2のコンピュータ・システムに供給される。変更は、一群のファイルの第1のコピーに対してなされる。エージェントは、一群のファイルの第1のコピーに対して供給される。各エージェントは、一群のファイルの第1のコピーにおけるファイルをアクセスしかつ読取るためのアクセス権を有する。リコンサイラ (reconciler) は、一群のファイルの第2のコピーを一群のファイルの第1

のコピーと一致させるために第2のコンピュータ・システムで供給される。プロキシ (proxy) は、一群のファイルの第1のコピーのエージェントからリコンサイラへ与えられる。プロキシは、一群のファイルの第1のコピーにおけるファイルをアクセスしかつ読取るために限定された権限 (limited authority) をリコンサイラに与える。次に、リコンサイラは、一群のファイルの第1のコピーに対してなされた変更が一群のファイルの第2のファイルにもなされるようにリコンサイラを用いて一群のファイルの第2のコピーを一群のファイルの第1にコピーと一致させる。本発明の最後の観点によれば、方法は、分散型システムにおいて実施される。この方法では、非均質ファイル・システム (heterogeneous file systems) が分散型システムに供給される。それに保持されるファイルへのアクセスを管理するために、記憶管理プログラム (storage manager) が各ファイル・システムに対して供給される。ファイルの第1のセットをファイルの第2のセットと一致させるための要求に応じて、ファイルの第1のセットを保持するファイル・システムに対する記憶管理プログラムによってアクセスがファイルの第1のセットへ与えられ、かつファイルの第2のセットを保持するファイル・システムに対する記憶管理プログラムによってアクセスがファイルの第2のセットへ与えられる。第1のオブジェクト・セットは、ファイルの第1のセット及びファイルの第2のセットを保持するそれぞれのファイル・システムの記憶管理プログラムの制御の下で第2のオブジェクト・セットと一致させられる。

【0017】
【実施例】本発明の好ましい実施例は、分散型環境 (distributed environment) で用いられるレプリケーション・ファシリティを提供する。レプリケーション・ファシリティは、システムの分散型ネーム空間における存続オブジェクト (persistent objects) のあらゆるツリーの弱一貫レプリケーション (weakly consistent replication) を支援する。レプリケーション・ファシリティは、単独オブジェクト (single objects) をレプリケート (模写、複製: replicate) する、または多重オブジェクトを含む論理的構造 (logical structures) をレプリケートする。レプリケーション・ファシリティは、オブジェクトの局所コピー (local copies) をオブジェクトの遠隔コピー (remote copies) と一致させる。リコンシリエーション (一致、reconciliation) は、オブジェクトの局所セットにおける各オブジェクトが局所オブジェクトの遠隔セットにおける (それに) 対応するオブジェクトと一致させられるように一対様式基準 (a pair-wise basis) で発生する。リコンシリエーションは、非均質ファイル・システム (heterogeneous file systems) にわたって発生しう。

【0018】図1は、本発明の好ましい実施例を実行す

るのに適当な分散型システム10を示す。分散型システム10は、多数の異なるデータ処理資源を相互接続する、ローカル・エリア・ネットワーク（LAN）、広域ネットワーク（WAN）、または他の相互接続機構のような、相互接続機構12を含む。データ処理資源は、ワークステーション14、16、18及び20、プリンタ22及び24、そして二次記憶装置26及び28を含む。ワークステーション14、16、18及び20のそれぞれは、各々メモリ30、32、34及び36を含む。メモリ30、32、34及び36のそれぞれは、分散型オペレーティング・システム38のコピーを保持する。各ワークステーション14、16、18及び20は、個別のファイル・システムを実行しうる。当業者は、本発明が図1に示す構成以外の構成で実施しようということに注目するであろう。図1に示す分散型システム10は、単に説明のために企図したものであり、本発明を限定するものではない。例えば、相互接続機構12は、個別のネットワーク・オペレーティング・システムを実行している多数のネットワークを一緒に相互接続しうる。

【0019】本発明の好ましい実施例は、ユーザ及びシステム管理責任者に存続“オブジェクト”をレプリケートさせる。このコンテキストにおいて、オブジェクトは、少なくとも一つのデータ・フィールド（data field）を保持する論理的構造である。類似な特性（similar properties）及び共通な意味（コモン・セマンティクス：common semantics）を有するオブジェクトの群は、オブジェクト・クラス（object classes）に編成される。多数の異なるオブジェクト・クラスが分散型システム10について規定されうる。本発明の好ましい実施例は、オブジェクトを採用するけれども、当業者は、本発明がオブジェクト指向型環境だけに限定されない；それよりも、本発明は、非オブジェクト指向型環境においても実施されうることには注目するであろう。本発明は、オブジェクトのレプリケーションに限定されない；それよりも、ファイルやファイル・ディレクトリーズ（file directories）のような、論理的構造のレプリケーションを支援することがより一般化される。オペレーティング・システム38は、本発明の好ましい実施例で用いられるオブジェクトを記憶するためのファイル・システムを含む。オブジェクトは、分散型ネーム空間19に編成される（図2参照）。分散型ネーム空間19は、オペレーティング・システム38のファイル・システムに記憶されたオブジェクト・ネーム21から形成された論理的ツリー式構造（logical tree-like structure）である。分散型ネーム空間19は、システム10の名前付きオブジェクト（named objects）の中で階層を示す（図1参照）。本発明の好ましい実施例のレプリケーション・ファシリティは、オブジェクトが分散型システムにわたって分散されうるようにオブジェクトのデュプリ

ケーション（複写：duplication）に対して供給するだけでなく、オブジェクトの多重コピーのリコンシリエーション（即ち、マルチマスター・レプリケーション：multimaster replication）に対しても供給する。リコンシリエーションは、オブジェクトが変更されたオブジェクトに対してなされた変更を反映するように、オブジェクトを変更されたオブジェクトとリコンサイルする（一致させる）ことを意味する。例えば、オブジェクトの遠隔コピーが変更されかつオブジェクトの局所コピーがその変更を反映すべくまだ更新されていないものと仮定する。各オブジェクトは、分散型ファイル・システム内にコンテンツ（内容）を有するだけでなくネーム（名前）及び位置をも有する。リコンシリエーションは、オブジェクトの局所コピーがオブジェクトの遠隔コピーが変更された方法と同様なファシリティで変更されるようにオブジェクトの二つのコピーを一致させることを含む。ここで用いられるように、項目“レプリケーション”は、オブジェクトの多重コピーが分散型システム10にわたって分散されるようにオブジェクトをデュプリケーション（複写する）ことだけを意味するのではなく、オブジェクトのコピーのリコンシリエーション（一致）をも意味する。

【0020】以下により詳細に本発明の好ましい実施例を説明する前に、以下で参照される2～3のキー・コンセプトを導入することは、有益である。“オブジェクト・セット（object set）”は、レプリケーションに対して一緒にグループ化されるオブジェクトのコレクション（collection）である。オブジェクト・セットは、単一のオブジェクトまたはオブジェクトのサブツリーを含むうる。オブジェクト・セットは、レプリケーションを要求するユーザまたは管理責任者によって特定される。

“レプリカ・セット（replica set）”は、対照的に、それぞれがオブジェクト・セットの局所コピーを所有するシステムのコレクションであり、“レプリカ（replica）”は、レプリカ・セットの構成部材（member）である。横たわっている物理的記憶システム（underlying physical storage system）（例えば、オブジェクトを記憶すべく採り入れられたファイル・システムの形式）からレプリケーション・ファシリティを防護（隔離）する（insulate）ためにかつ拡張性を供給するために、本発明の好ましい実施例は、レプリケートされたオブジェクト記憶のアブストラクション（abstraction：抽象）

（RepIStore）を採用する。RepIStore・アブストラクションは、レプリケーション・ファシリティに非均質ファイル・システムにわたって印加されることを許容する。RepIStoreは、レプリケーション・ファシリティーズを支援すべく横たわっている物理的記憶システムに対して支援されなければならない一群のインタフェースを表わす。特に、RepIStoreインタフェースを支援するオブジェクト記憶（obje

ct stores) に存在するこれらのオブジェクトだけがレプリケートされうる。インタフェースは、論理的に関連した機能の一名前付き群 (a named group) である。インタフェースは、インタフェースによって供給された一群の関連機能に対する (パラメータのような) シグネチャ (signatures) を特定する。インタフェースは、機能を実行するコード (符号) を供給しない; それよりも、機能を実行するためのコードがオブジェクトによってまたは他の実施 (other implementations) によって供給される。インタフェースの瞬間に対してコードを供給するオブジェクトは、インタフェースを“支援する”と言われる。インタフェースを支援するオブジェクトによって供給されるコードは、インタフェース内で特定されたシグネチャに従わなければならない。それゆえに、上述した実施例では、オブジェクト・セットでオブジェクトを記憶するオブジェクト記憶は、オブジェクト・セットがレプリケートされるために Rep l Store インタフェースを支援しなければならない。Rep l Store インタフェースの実行 (実施) は、ファイル・システムのそれぞれにわたってレプリケーションを支援するために分散型システム 10 内のファイル・システムのそれぞれに対して供給される。

【0021】各 Rep l Store は、局所ボリューム上のレプリケートされたオブジェクトを識別するための機構を供給する。この機構は、レプリケートィド・オブジェクト (replicated object) ID (ROB ID) である。ROB ID は、識別及びレプリケートされているオブジェクトについての他の情報をエンキャプスレート (encapsulate : 密閉) するアブストラクションである。Rep l Store は、ROB IDs を直列化し (serializing) かつ非直列化する (deserializing) ためのルーチンを支援する。オブジェクトの ROB ID は、多数の動作を実行する機構を供給する。例えば、オブジェクトは、ROB ID に包含された情報を用いて記憶装置から検索されうる。更に、オブジェクトの構成要素ネーム (component name) は、その ROB ID から導出されうる。各 Rep l Store は、レプリケートィド記憶変更記録簿 (replicated storage change log) 40 を維持する (図 3 参照)。変更記録簿 40 は、オブジェクト・セットのオブジェクトに対してなされた変更を特定する多数の変更項目 42 を含む。各変更項目 42 は、形式フィールド (type field) 44、変更されたオブジェクトに対する直列化された ROB ID フィールド 46、変更が発生した時間 (局所時間) を示す時間フィールド (time field) 48、及び変更に対応付けられた RIB を保持するレプリケーション情報ブロック (RIB) フィールド 50 を含む。ここの述べられた実施例では、形式フィールド 44 内で特定されうる五つの形式の変更が存在する。これらの変更は、消去 (deletion)、生成 (creation)、変更 (modification)、改名 (rena

ming)、及び移動 (moving) である。消去は、オブジェクトが消去されるときに発生する。生成は、オブジェクトが生成されるときに発生する。変更は、オブジェクトの内容 (contents) がある方法で変更されるときに発生する。改名は、オブジェクトの構成要素ネームが変更されるときに発生し、そして移動は、オブジェクトがシステムの分散型ネーム空間の新しいペアレント (new parent) の下で移動されるときに発生する。

【0022】カーソル 49 は、変更項目 42 のリストへのインデックス (索引、指標) として作用する変更記録簿 40 内で維持される。カーソル 49 は、変更項目 42 のリストでマーカ (marker) として作用する。加えて、変更記録簿は、多重カーソルを含む。カーソル 49 は、タイム・スタンプ (時刻表示 : time stamp) の形を取りうる。カーソル 49 は、例えば、ある時点の後で発生した変更の開始を識別しうる。レプリケートされているオブジェクト・セットのあらゆるオブジェクトは、RIB 51 でスタンプされる (図 4 参照)。RIB 51 は、三つのフィールドを有する : オリジネータ・フィールド (originator field) 53、変更識別子フィールド (change identifier field) 55、及びプロパゲータ・フィールド (propagator field) 57。オリジネータ・フィールド 53 は、オブジェクトへの最後の変更が発生したところを特定する。変更識別子フィールド 55 は、対照的に、オリジネータ・フィールド 53 内で識別されたオリジネータに関するオブジェクトへの最後の変更を識別する。最後に、プロパゲータ・フィールド 57 は、局所側へ変更を送ったパーティの識別を特定する。オブジェクトが論理的に変更されるときには、オブジェクトに対応付けられた RIB 51 は、オリジネータ及びプロパゲータとして局所側を反映すべく変更される。変更識別子は、適当にスタンプされる。

【0023】レプリケーションは、負荷平衡 (load balancing) 及び可用性 (availability) を供給することにおいて分散型システム 10 に対して有用である。レプリケーションは、オブジェクトのいずれか一つのコピー上に負荷を制限すべく分散型システム 10 にわたって記憶されたオブジェクトの一つ以上のコピーを有することによって負荷平衡を供給する。レプリケーションは、重要なオブジェクトの多重コピーをシステム 10 にわたって分散されることを許容することによって可用性を向上する。向上した可用性は、システムの障害許容力を増大する。特に、システム 10 にわたって分散された重要なオブジェクトのコピーを有することによって、ユーザは、オブジェクトへのアクセスを妨げるかまたは制限するシステム内の故障によって影響されなくなる。向上した可用性は、システムの性能をも向上する。本発明の好ましい実施例は、オペレーティング・システム 38 の部分であるレプリケーション・ファシリティ 54 (図 5 参照) に具体的に表されている。それにもかかわらず、当業者

は、本発明のレプリケーション・ファシリティが、図形ユーザインタフェースを含んでいる、他の環境においても実行されうること注目するであろう。図5に示すように、レプリケーション・ファシリティ42は、三つの主機能構成要素を含んでいる：複写構成要素56、リコンサイラ構成要素58及び制御構成要素60。レプリケーション・ファシリティ54は、デублиケーションに対して複写構成要素56を用いる。加えて、レプリケーション・ファシリティ54は、それらが互いに一貫していることを確実にすべくリコンサイラ構成要素58を用いてオブジェクト・セットのコピーを一致させる。このリコンサイレーションは、分散型システム10にわたるオブジェクトの一貫した視点 (consistent view) を保証する。

【0024】制御構成要素56により及ぼされた (exerted) 制御の一つのレベルは、レプリケーションが呼び出された方法に関する。レプリケーションは、手動的または自動的に呼び出されうる。手動呼出しは、レプリケートすべく明示的な要求がユーザまたは他のパーティによってなされることを要求する。ユーザまたは他のパーティは、オブジェクト・セット及びレプリケーションの宛先を特定しなければならない。宛先は、各レプリケーション・サイクルに対して特定されない：それよりも、レプリカ接続 (replica connection) が最初に特定される。レプリカ接続は、レプリケーションに含まれるべき二つのレプリカとオブジェクト・セットを識別する。対照的に、自動呼出しは、ある一定の事象67 (図6参照) または (事象の形式として解釈されうる) ある一定時間の経過によって始動 (トリガー) される。レプリケーションは、固定された時間間隔で発生すべく予め計画されうる。制御機構により及ぼされる制御の別の観点、レプリケーションを呼び出しうるものに関する。レプリケーションは、適正特権パーティ (appropriately privileged party) によって呼び出されうる。本発明の好ましい実施例は、二つの形式のレプリケーションを供給する：公開レプリケーション及び専用レプリケーション。公開レプリケーションは、オブジェクトの“公開”コピーを生成すべく適正特権パーティズによってのみ実行されうる処理を意味する。公開レプリケーションでは、生成されるオブジェクト・セットのコピーのそれぞれは、一貫性を維持すべく他のコピーと協同 (協働) する。公開コピーを記憶するネーム空間のノードは、全体として、公開レプリカ・セットを形成し、そしてセットのメンバ (members) は、コピーの間で一貫性を維持すべく状態情報 (state information) を保つ。オブジェクト上のアクセス制限は、保存される。オブジェクト・セットの公開コピーで発生する変更は、他の公開コピーと一致させられる。

【0025】専用レプリケーションは、オブジェクト・セットの専用コピーを生成する処理を意味する。専用コ

ピーは、非管理責任者を含む、あらゆるパーティによって生成されうる。レプリケートド・セットの全てのメンバが、コピーの間の一貫性を維持すべく状態情報を保つのではない。専用レプリケーションは、以下に詳細に説明する。多数の構成要素は、本発明の好ましい実施例におけるレプリケーション処理で役割をする。図6は、公開レプリケーション処理で重要な役割をしうる構成要素を示す図である。オブジェクト・レプリケーション・エージェント (ORAs) 62及び64は、レプリケーションのための自動支援 (automatic support) を供給すべくオブジェクト・セットが記憶されるノードのためにエージェントとして作用するレプリケータ・オブジェクト (replicator objects) である。分散型システムの各機械は、それ自体のORAを有する。ORAs 62及び64は、遠隔クライアントのためになされた要求を提供する遠隔手続き呼出しサーバ (remote procedure call (RPC) servers) として作用しうるかまたは類似な役割を提供する他の形式の高信頼通信機構 (reliable communication mechanisms) で代替的にありうる。個別のORA 64 (62) は、局所オブジェクトに対して提供され、別のORA 64は、公開レプリケーション処理の対応する遠隔オブジェクトに対して提供される。局所ORA 62は、Rep1Store DLL 66及びRep1Store Manager DLL 65のローディングを司る。Rep1Store Manager 65は、Rep1Store 66へのアクセスの調整を司る。クライアントは、所与の物理的記憶システムに対して適当なRep1Store 66をロードすべくRep1Store Manager 65を呼出す。ORAs 62及び64は、それらに局所オブジェクト記憶 (装置) からレプリケートされている全てのオブジェクトを読取らせかつ書込ませる特権のレベルを有する。ORAs 62及び64は、公開レプリカ (public replicas) を維持する他のORAsと変更を交換すべく要求に応答することを司る。

【0026】リコンサイラ68も公開レプリケーション処理で役割をする。それは、局所オブジェクト・セットを対応する遠隔オブジェクトと一致させるべく局所ORA 62に対するカウンターパート (counterpart) として作用する。リコンサイラ68は、局所ORA 62によって呼び出されかつリコンサイル (一致) されるべきオブジェクトを開くことを司る。二つの形式のリコンサイラ・オブジェクトがリコンサイラ68によって呼び出されうる。特に、クラス特定リコンサイラ70またはデフォルト (即ち、クラス独立: class-independent) リコンサイラ72が呼び出されうる。クラス特定リコンサイラ70は、レプリケーションでクラス特定要求を有するオブジェクトを一致させる。クラス特定リコンサイラ70は、一クラスのオブジェクトに対してのみ適用される。クラス独立リコンサイラ72は、それらのクラス

の如何に係わらずオブジェクトを一致させる。多重クラス独立リコンサイラがシステム 10 で利用可能でありうる。例えば、各オブジェクト・セットは、それ自体のクラス独立リコンサイラを有しうる。全てのレプリカ・セットは、クラス特定リコンサイラが利用不可能であるときにはいつでも呼び出されるそれ自体のクラス独立リコンサイラに対応付けられうる。最後に、上述したように、事象 67 は、レプリケーションをトリガ（始動）する役割をしうる。

【0027】図 7 は、本発明の好ましい実施例におけるレプリケーションに対して実行されるステップのフローチャートである。最初に、遠隔オブジェクト・セットに対する変更記録簿 40（図 3 参照）に対してアクセスが獲得される（図 7 のステップ 74）。特に、局所オブジェクト・セットが遠隔オブジェクト・セットと一致させられるときには、局所 ORA 62（図 6 参照）は、遠隔手続き呼出し機構（remote procedure call mechanism）を介して遠隔 ORA 64 と接触する。局所 ORA 62 は、変更記録簿 40 へのアクセスを獲得すべく遠隔 ORA 64 と接触する。次に、カーソル 49（図 3 参照）は、変更記録簿で生成される（図 7 のステップ 76）。特に、局所 ORA 62 は、オブジェクト・セット間の最後のリコンシリエーションの時間を示しているタイム・スタンプを記憶し、そして次にカーソル 49 として用いられるべくこのタイム・スタンプを遠隔 ORA 64 へ渡す。次に、遠隔 ORA 64 は、このタイム・スタンプをカーソルとして遠隔変更記録簿 40 へ渡す。カーソルは、最後のリコンシリエーションの後でタイム・スタンプを有する変更記録簿の項目を識別し、それゆえに、（項目は、）このレプリケーション・サイクルに対して関係がある。

【0028】そして、変更項目の一覧表（リスト）は、最後のリコンシリエーションの後で発生した変更に対する変更項目を識別すべく、カーソルを用いて遠隔変更記録簿から得られる。遠隔 ORA 64 は、遠隔 ORA が局所 ORA で発生した変更を局所 ORA 62 へパス・バックしない（即ち、遠隔 ORA が RIBs のオリジネータ・フィールド 53 を検査する）ことを保証すべく変更項目 42 のそれぞれの RIBs 51 を予備選択（screen）し、かつ局所 ORA から伝送された変更に対する変更項目（即ち、遠隔 ORA が RIBs のプロパゲータ・フィールド 57 を検査する）が送られないことを保証すべく RIBs を検査する。結果として得られた（合成）変更項目は、それらが持続的に記憶される局所 ORA 62 へパス・バックされる。局所 ORA 62 は、変更項目の RIBs によって識別されたオブジェクト上でネーム空間リコンシリエーション（ステップ 80）及び内容リコンシリエーション（ステップ 82）を実行すべくリコンサイラ 68 を用いる。特に、リコンサイラ 56 は、遠隔オブジェクト・セットで変更した各オブジェクトを局

所オブジェクト・セットの対応するオブジェクトと一致させる。遠隔オブジェクトに対してなされたあらゆる変更は、対応する局所オブジェクトに対してなされる。クラス特定リコンサイラ 70 またはクラス独立リコンサイラ 72 が用いられるかどうかは、ソース（即ち、オブジェクトの遠隔コピー）に依存する。クラス特定リコンサイラ 70 は、もしオブジェクトの遠隔コピーがそのようなリコンサイラを要求するときのみ用いられる。

【0029】ネーム空間リコンシリエーションは、内容変更に限定されないかまたはシステム特性に対応付けられない変更項目に記録されたあらゆる変更に対して実行される。そのような変更は、生成、消去、移動、及び改名を含む。ネーム空間リコンシリエーションは、対応している遠隔オブジェクトに対して記憶された情報に関して局所オブジェクトの RIBs によって取得可能な情報を比較することによって発生する。ネーム・レゾリューション・コンフリクト（ネーム解像矛盾：name resolution conflicts）を解決するための多くの異なる方法が本発明内で用いられうる。本発明の好ましい実施例は、しかしながら、規則（rules）を採用する。ネーム空間矛盾（namespace conflicts）を解決すべく本発明の好ましい実施例によって用いられる第 1 の規則は、先の変更にはわたる最後の変更を選択することである。第 1 のネームを有すべく一つの側でオブジェクトが移動／改名され、そして異なるネームを有すべく同じオブジェクトが別の側へ移動または改名されるときに、オブジェクトが最後の変更に対応付けられたネームを仮定（想定）するように最後に発生される変更が選ばれる。第 2 の規則は、ネーム空間衝突を解決すべく用いられる。ネーム空間衝突は、二つの異なるオブジェクトが生成され、移動され、または同じネームを有すべく改名されるときに発生する。第 2 の規則は、オブジェクトが生成され、移動され、または改名されたうちのいずれであっても、最初は、局所側のオブジェクトに対して選択されたネームであることを特定する。

【0030】内容リコンシリエーション（図 5 のステップ 70 を参照）は、局所オブジェクトが遠隔オブジェクトに対してなされた変更を含むように局所オブジェクトの内容を遠隔オブジェクトに一致させることを含む。変更記録簿の変更を検査することによって、局所オブジェクトは、遠隔オブジェクトと同じ内容を有すべく変更されうる。レプリケーション中、変更は、一つのレプリカから別のものへ伝搬される。レプリケーションは、オブジェクト・セットの最初のコピーに対してなされた変更がオブジェクト・セットの第 2 のコピーになされるということにおいて“一方通行”である。オブジェクト・セットの第 2 のコピーに対してなされた変更をオブジェクト・セットの第 1 のコピーへ複写するための即時反作用（immediate reciprocal action）が存在しない。しかしながら、オブジェクト・セットの第 1 のコピーへのそ

のような伝搬は、実行されうる。このレプリケーションの一方通行性質を与えて、各レプリカは、レプリカに対してオブジェクト・セットの局所コピーがどのように更新するかを監視し、カーソルは、共同変更記録簿に維持される。リコンシリエーション中の各交換の完了で、二つのレプリカは、カーソル情報を交換する。

【0031】公開レプリケーションは、多数の機密保護イシュー（発行：issues）をポーズ（主張：poses）する。一般に、リコンサイラは、レプリケーションを実行するためにオブジェクトを更新できなければならない。クラス独立リコンサイラは、信用システム処理であり、それゆえに、機密保護リスクをポーズしない。しかしながら、クラス特定リコンサイラは、信用システム処理ではない、それゆえに機密保護スレット（脅威：threat）をポーズする。この機密保護ジレンマを緩和することを援助すべく、本発明の好ましい実施例は、“プロキシ（proxies）”を利用する。プロキシ（proxy）は、異常な特権（extraordinary privileges）を有することなく明確な動作を実行するワーカー処理（worker processes）または遠隔処理を許容するデリゲーション・チケット（delegation ticket）である。プロキシは、許可しているパーティ（granting party）の資格（credentials）をパッケージ（package）しかつ遠隔オブジェクトへのアクセスを求めているパーティにそれらを貸す。アクセスを求めているパーティは、許可しているパーティに取って換わりかつ必要なオブジェクトにアクセスする。これらの資格は、暗号化されうる。図8は、本発明の好ましい実施例においてプロキシを利用すべく実行されるステップのフローチャートである。レプリケーション処理中、遠隔ORA64（図6）は、局所リコンサイラ68にプロキシを与える（図8のステップ84）。上述したように、このプロキシは、局所リコンサイラへ遠隔ORAによって許可されるべきアクセス権及び適当な資格を含む。リコンサイラ68は、遠隔側へ資格を送る（図8のステップ86）。換言すると、リコンサイラ68は、遠隔側へプロキシを贈る。遠隔側は、資格を検査し、そしてもし資格が有効であるならば、問題のオブジェクト・セットの遠隔コピー内のオブジェクトへの限定付きアクセスを許可する（ステップ88）。リコンサイラ68は、オブジェクト・セットの遠隔オブジェクトへのアクセスを獲得する（ステップ80）。しかしながら、局所リコンサイラのアクセス範囲は、相応なリコンシリエーションを実行するために必要なものだけに限定される。本発明は、排他的にプロキシを用いることに限定されないということに注目すべきである。各ORAをアクセス権を許可する共通アクセス群のメンバーにするような、セキュア・アクセス（secure access）を許可するあらゆる技法が許容可能である。

【0032】上記説明の大部分は、公開レプリケーションに集中している。専用レプリケーションは、公開レ

リケーションに類似であるが、しかし多数の相違点を含む。専用レプリケーションでは、変更の源（ソース）は、どのオブジェクトがデュプリケート（複写）または変更されたかという記録を維持しない。ソースで維持される状態情報が存在しない。ソースは、変更が発生したことを忠告することを司らない。従って、公開レプリケーションに対して要求される資源は、要求されない。これらの特徴は、レプリケーションの手動制御が望ましい場合、またはオブジェクト・セットの公開コピーを維持する経費が保証されない場合に専用レプリケーションを特に適切にする。本発明は、その好ましい実施例を参照して記述されたが、当業者は、添付した特許請求の範囲に規定される本発明の範囲から逸脱することなく、形式及び詳細における種々の変更がなされうることについて注目するであろう。例えば、本発明は、オブジェクト指向型環境にて実行される必要がなく、かつ図1に示すような分散型システム構成で単独で実施される必要もない。更に、RPC機構以外の通信機構が、遠隔相互作用に対して用いられうるし、プロキシ以外の機密保護機構が採用されうる。

【0033】

【発明の効果】本発明の方法は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含む多数のコンピュータ・システムを有している分散型システムにおいて、ファイルのツリーに編成された複数のファイルを提供し、ファイルのコピーが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて一つのコンピュータ・システムの記憶装置に記憶される単一のファイルをレプリケートし、ファイルのサブツリーのコピーが別のコンピュータ・システムの記憶装置に記憶されるようにレプリケーション・ファシリティを用いて一つのコンピュータ・システムの記憶装置に記憶される、ファイルのツリーから、多重段階のファイルのサブツリーをレプリケートする段階を具備するので、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給し、かつ多重マスターの弱一貫性のレプリケーションを供給することができる。本発明の分散型システムは、それぞれがファイルを記憶する記憶装置を含んでいる複数のコンピュータ・システムと、ファイルのネームのツリー構造としてシステムのネーム空間を管理するネーム空間管理プログラムと、ネーム空間のサブツリーを含んでいるあらゆるファイルをレプリケートするレプリケーション・ファシリティとを備えているので、分散型環境の分散型ネーム空間内であらゆるサブツリーをレプリケートすることが可能である。

【0034】本発明の方法は、リコンサイラ・ファシリティ及び多数のコンピュータ・システムを有している分散型システムにおいて、一つのコンピュータ・システムのファイルの第1のコピー及び別のコンピュータ・シ

テムのファイルの第2のコピーを供給し、最後に一致されて以来ファイルの第1のコピーに対してなされたあらゆる変更をファイルの第2のコピーが組み込むようにリコンサイラ・ファシリティを用いてファイルの第1のコピーをファイルの第2のコピーと一致させ、一つのコンピュータ・システムの一組のファイルの第1のコピー及び別のコンピュータ・システムの一組のファイルの第2のコピーを供給し、最後に一致されて以来一組のファイルの第1のコピーに対してなされたあらゆる変更を一組のファイルの第2のコピーが組み込むようにリコンサイラ・ファシリティを用いて一組のファイルの第1のコピーを一組のファイルの第2のコピーと一致させる段階を具備するので、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給する。本発明の方法は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含んでいる多数のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムの記憶装置に記憶される一組のファイルの第1のコピーを供給し、第2のコンピュータ・システムの記憶装置に記憶される一組のファイルの第2のコピーを供給し、一組のファイルの第1のコピーのファイルの少なくとも一つに対して変更を行い、事象の発生により一組のファイルの第2のコピーへ変更を伝搬し、一組のファイルの第1のコピーのファイルの少なくとも一つに対して追加の変更を行い、別の事象の発生により一組のファイルの第2のコピーへ追加の変更を伝搬する段階を具備するので、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給する。

【0035】本発明の方法は、レプリケーション・ファシリティ及びそれぞれが記憶装置を含むコンピュータ・システムを有している分散型システムにおいて、コンピュータ・システムの記憶装置に、ネームを有している、ファイルを記憶し、記憶されたファイルのネームの論理的編成からなる分散型ネーム空間を供給し、ファイルの選択された部分を保持している新しいファイルを生成するためにレプリケーション・ファシリティを用いてそのネームが分散型ネーム空間の一部を形成しかつ一つのコンピュータ・システムの記憶装置に記憶された一組のファイルの選択された部分をレプリケートする段階を具備するので、分散型環境の分散型ネーム空間内であらゆるサブツリーをレプリケートすることが可能である。本発明の方法は、第1のコンピュータ・システム及び第2のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムに記憶される所与のクラスの一組のファイルの第1のコピーを供給し、第2のコンピュータ・システムに記憶される所与のクラスの一組のファイルの第2のコピーを供給し、所与のクラスのファイルだけを一致させるクラス特定リコンサイラを用いて一組のファイルの第1のコピーを一組のファイルの第2のコピーと一致させる段階を具備するので、

分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給する。

【0036】本発明の方法は、専用レプリケーション機構及びそれぞれが記憶装置を含んでおり処理を実行するためのコンピュータ・システムを有している分散型システムにおいて、一つのコンピュータ・システム上でアプリケーション・プログラムを実行し、アプリケーション内にてそれぞれがファイルにアクセスすることを許容される処理のリストを維持している一組のファイルをレプリケートすべく専用レプリケーション機構に対して要求を行い、ファイルをアクセスすることを許容される処理のリストを、各ファイルに対して、レプリケートすることなく新しい組のファイルを生成すべく専用レプリケーション機構を用いて一組のファイルをレプリケートする段階を具備するので、多重マスターの、弱一貫性のレプリケーションを供給し、公開レプリケーションと専用レプリケーションの両方を支援するという効果を奏する。本発明の方法は、第1のコンピュータ・システム及び第2のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムでファイルのコレクションを供給し、ファイルのコレクションを第2のコンピュータ・システムへレプリケートするための要求に応じて、コレクションの全てのファイルがレプリケートされるべきかまたはされるべきでないかを決定し、コレクションの全てのファイルがレプリケートされるべきであると決定されるならば、コレクションのレプリカが第2のコンピュータ・システムで供給されるようにコレクションの全てのファイルをレプリケートし、コレクションの全てのファイルがレプリケートされるべきでないと決定されるならば、コレクションのファイルの全てをレプリケートしない段階を具備するので、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給し、多重マスターの、弱一貫性のレプリケーションを供給する。

【0037】本発明の方法は、第1のコンピュータ・システム及び第2のコンピュータ・システムを有している分散型システムにおいて、第1のコンピュータ・システムの一組のファイルの第1のコピーを供給し、第2のコンピュータ・システムの一組のファイルの第2のコピーを供給し、一組のファイルの第1のコピーに対して変更を行い、それぞれが一組のファイルの第1のコピーのファイルをアクセスしかつ読取るためのアクセス権を有するエージェントを一組のファイルの第1のコピーに対して供給し、一組のファイルの第2のコピーを一組のファイルの第1のコピーと一致させるために第2のコンピュータ・システムでリコンサイラを供給し、一組のファイルの第1のコピーのファイルをアクセスしかつ読取るべく限定された権限をリコンサイラに賦与するプロクシを、一組のファイルの第1のコピーのエージェントからリコンサイラへ賦与し、一組のファイルの第1のコピー

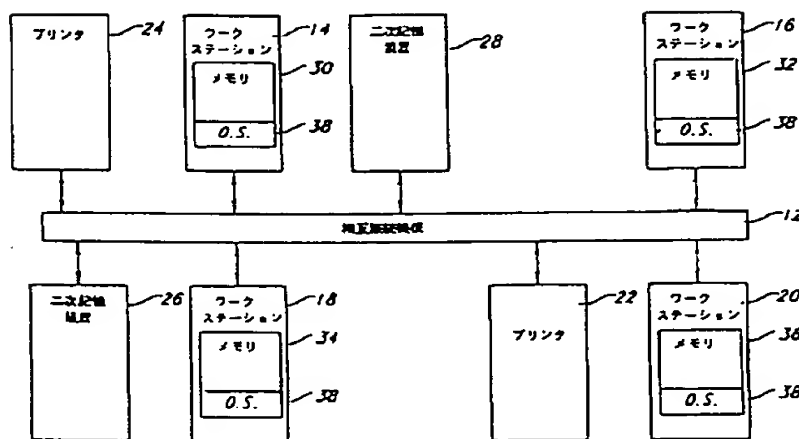
25

に対してなされた変更が一群のファイルの第2のコピーに対してなされるようにリコンサイラを用いて一群のファイルの第2のコピーを一群のファイルの第1のコピーと一致させる段階を具備するので、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給し、多重マスターの、弱一貫性のレプリケーションを供給する。

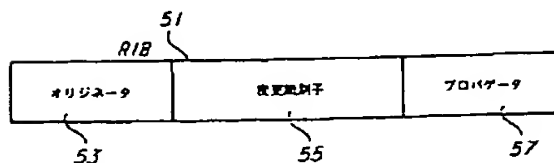
【0038】本発明の方法は、分散型システムにおいて、分散型システムの非均質ファイル・システムを供給し、ファイル・システムのファイルへのアクセスを管理すべく各ファイル・システムに対して記憶管理プログラムを供給し、第1の組のファイルを第2の組のファイルと一致させるための要求に応じて、第1の組のファイルを保持するファイル・システムに対する記憶管理プログラムにより第1の組のファイルへのアクセスを賦与し、かつ第2の組のファイルを保持するファイル・システムに対する記憶管理プログラムにより第2の組のファイルへのアクセスを賦与し、第1の組のファイル及び第2の組のファイルを保持しているそれぞれのファイル・システムの記憶管理プログラムの制御の下で第1のオブジェクト・セットを第2の組のオブジェクトと一致させる段階を具備するので、分散型環境におけるファイルまたはファイルの部分のレプリケーションを供給し、多重マスターの、弱一貫性のレプリケーションを供給する。

【図面の簡単な説明】

【図1】



【図4】



26

【図1】本発明の好ましい実施例を実施するのに適当な分散型システムのブロック図である。

【図2】本発明の好ましい実施例による分散型システムに対する分散型ネーム空間の図である。

【図3】本発明の好ましい実施例に用いられる変更記録簿のブロック図である。

【図4】本発明の好ましい実施例に用いられるレプリケーション情報ブロック (RIB) のブロック図である。

【図5】本発明の好ましい実施例に用いられるレプリケーション・ファシリティの機能構成要素 (functional components) を示しているブロック図である。

【図6】本発明の好ましい実施例における公開レプリケーション (public replication) で役割をする構成要素の相互作用を示す図である。

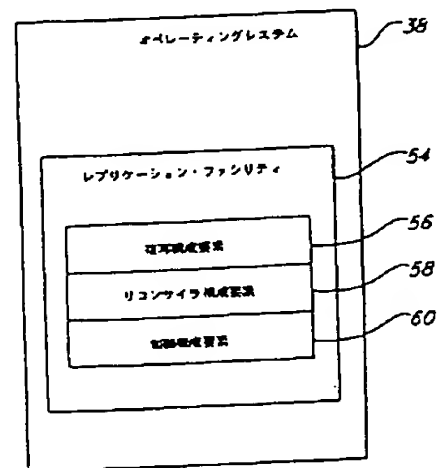
【図7】本発明の好ましい実施例におけるレプリケーションで実行されるステップのフローチャートである。

【図8】本発明の好ましい実施例におけるレプリケーションの間に機密保護を供給すべく実行されるステップを示すフローチャートである。

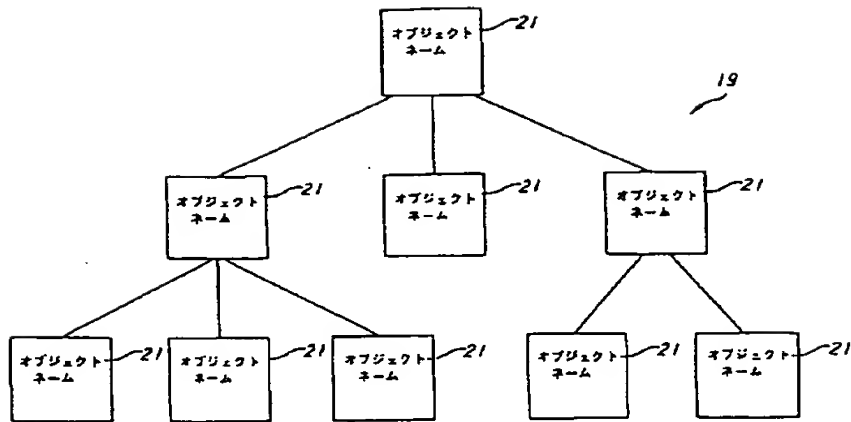
【符号の説明】

- 38 オペレーティング・システム
- 54 レプリケーション・ファシリティ
- 56 複写構成要素
- 58 リコンサイラ構成要素
- 60 制御構成要素

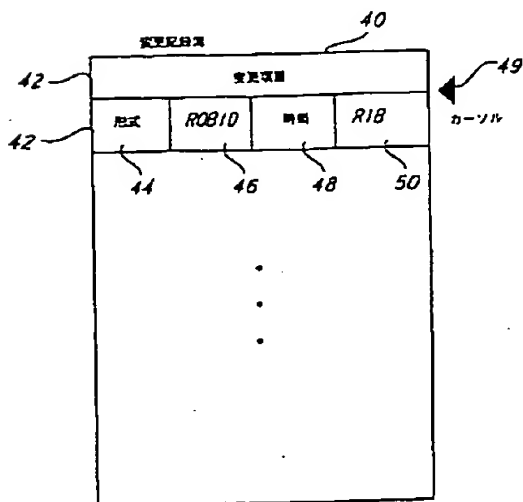
【図5】



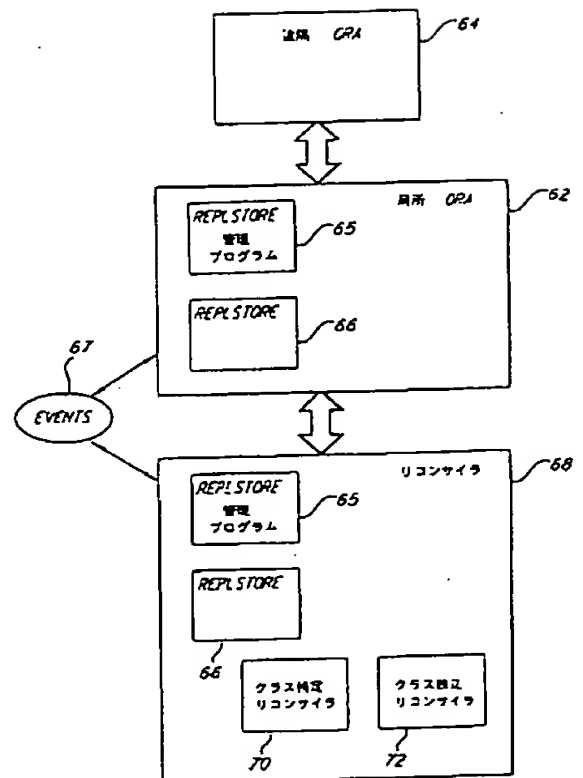
【図 2】



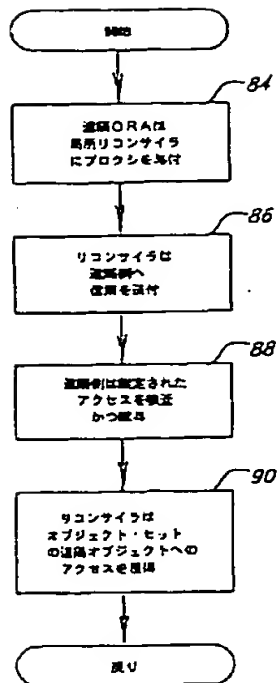
【図 3】



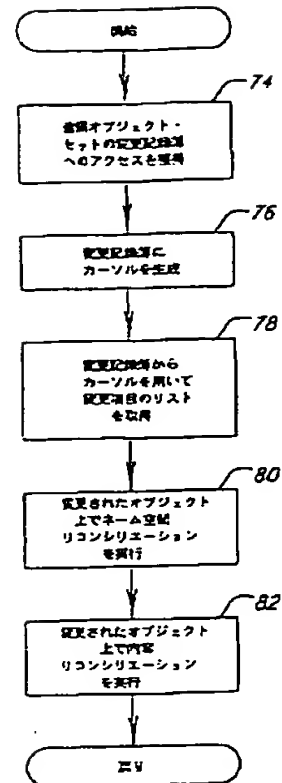
【図 6】



【図7】



【図8】



フロントページの続き

(72)発明者 ビーター ジェイ クック
アメリカ合衆国 ワシントン州 98027
イサカ シカモア ドライブ サウスイ
スト 1430

(72)発明者 アーノルド エス ミラー
アメリカ合衆国 ワシントン州 98005
ベルヴィュー サウスイースト トウエン
ティセカンド プレイス 12806
(72)発明者 バラン エス ラーマン
アメリカ合衆国 ワシントン州 98052
レッドモンド ワンハンドレッドアンドフ
ォーティセヴンス プレイス ノースイ
スト 6120